Color standard display function: A proposed extension of DICOM GSDF

Tom Kimpe,[a] Johan Rostang, Gert Van Hoey, and Albert Xthona
Barco NV, Healthcare Division, Beneluxpark 21, 8500 Kortrijk, Belgium

(Received 1 February 2016; revised 6 July 2016; accepted for publication 12 July 2016; published 4 August 2016)

Purpose: Color images are being used more in medical imaging for a broad range of modalities and applications. While in the past, color was mostly used for annotations, today color is also widely being used for diagnostic purposes. Surprisingly enough, there is no agreed upon standard yet that describes how color medical images need to be visualized and how calibration and quality assurance of color medical displays need to be performed. This paper proposes color standard display function (CSDF) which is an extension of the DICOM GSDF standard toward color. CSDF defines how color medical displays need to be calibrated and how QA can be performed to obtain perceptually linear behavior not only for grayscale but also for color.
Methods: The proposed CSDF algorithm uses DICOM GSDF calibration as a starting point and subsequently uses a color visual difference metric to redistribute colors in order to obtain perceptual linearity not only for the grayscale behavior but also for the color behavior. A clear calibration and quality assurance algorithm is defined and is validated on a wide range of different display systems.
Results: A detailed description of the proposed CSDF calibration and quality assurance algorithms is provided. These algorithms have been tested extensively on three types of display systems: consumer displays, professional displays, and medical grade displays. Test results are reported both for the calibration algorithm as well as for the quantitative and visual quality assurance methods. The tests confirm that the described algorithm generates consistent results and is able to increase perceptual linearity for color and grayscale visualization. Moreover the proposed algorithms are working well on a wide range of display systems.
Conclusions: CSDF has been proposed as an extension of the DICOM GSDF standard toward color. Calibration and QA algorithms for CSDF have been described in detail. The proposed algorithms have been tested on several types of display systems and the results confirm that CSDF largely increases the perceptual linearity of visualized colors, while at the same time remaining compliant with DICOM GSDF. *© 2016 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).*
[http://dx.doi.org/10.1118/1.4959544]

Key words: color, medical display, calibration, QA, medical imaging, DICOM GSDF, CSDF

1. INTRODUCTION

Consistent and high quality visualization of medical images are important. Medical images often contain subtle image features that are crucial for a radiologist in order to come to a correct diagnosis. It is important that radiologists can rely on the fact that medical images are accurately and consistently visualized independently of the specific display system that is being used. Although this may seem obvious, it is not obvious at all. Display systems are inherently unstable. Medical images shown on uncalibrated display systems will look different and there is no guarantee that subtle image features visible at a certain moment in time will be visible in the future as well, even on one and the same display system.[1,2]

DICOM has proposed a standard for the purpose of visualization of grayscale radiology images which ensures consistency over different devices and over time. DICOM GSDF is a relative calibration aiming to linearize the perception of luminance of a display without reducing its luminance dynamic range. The perceived variation is expressed as just noticeable difference (JND) and is based on Barten's contrast sensitivity function.[3]

A good introduction to the GSDF is proposed by Fetterly *et al.*[4] This standard display calibration is applicable for any grayscale medical imaging modality and it has been shown to have a positive impact on diagnostic performances.[5]

When all medical displays are DICOM GSDF calibrated, which is *de facto* the case today, applications can assume that any grayscale image they generate will be visualized by the display system in a perceptually linear way. DICOM GSDF is limited to grayscale and does not specify how color medical images need to be visualized. So when medical viewing applications generate color information, they cannot make the same assumption today, as we will discuss in this paper. The proposed color standard display function (CSDF) calibration has the goal to solve this problem to a large degree for the most commonly used color scales.

The present document intends to propose and describe a new method for color calibration. An initial validation of the proposed method has been performed and is reported. A full

quantification of its positive effect on diagnosis by means of observer studies will be the topic of a future paper.

2. PROBLEM STATEMENT

2.A. Consistent reproduction of colors

Color images are being used more in medical imaging for a broad range of modalities and applications. While in the past, color was mostly used for annotations and metadata, today color is also widely being used for diagnostic purposes. A few examples are the use of color in Doppler ultrasound, in contrast enhanced MRI of, e.g., the breast or prostate, in fused multimodality images, and for dermatology or ophthalmology applications.

For grayscale medical imaging, the widely adopted DICOM GSDF standard clearly describes how grayscale medical images need to be visualized to ensure consistent and high quality visualization and to maximize probability that small subtle image features will be visible. Surprisingly enough, there is no agreed upon standard yet that describes how color medical images need to be visualized and how calibration and quality assurance of color medical displays need to be performed. Two important questions are as follows: "Are such color calibration and quality assurance needed or not?"; and "is there significant variability and instability in the color behavior of display systems?"

There has been significant work done already over the past few years to characterize and describe the behavior of color display systems. One particularly relevant publication describes a consensus opinion of various stakeholders from the medical imaging community.[1] This work lists various forms of color instability in modern state-of-the-art display systems. A few examples are primary instability, variability of the white point, color gamut variability, neutral gray imbalance, instability of the color behavior over time (aging of the display), etc. Moreover it has been shown that there are large differences in color behavior between different types and brands of displays. All these instabilities are not compensated for by means of DICOM GSDF calibration since GSDF only calibrates the grayscale behavior of display systems and leaves the color behavior unchanged.

It is without doubt that such inconsistent color behavior is not desired nor acceptable for medical imaging applications. Imagine a radiology reading room with several workstations and display systems. Significant differences in color behavior of these displays could result in a situation where color medical images look different on the different displays in the reading room. Moreover in such a situation, it cannot be excluded that on some of these displays, subtle color features of the medical images will be visible while these same features are invisible on the other displays in that same reading room. This can be especially the case if one or more of these displays are replaced by new systems while some older systems are still being used. And if color behavior of these displays is not stabilized, then colors will gradually be visualized in a different way as the display ages. This could result in a situation where prior images that are being reviewed look differently than when they were first read in the past.

2.B. Color requirements in medical imaging

Depending on the specific field of medicine, requirements for the representation of colors may vary. For instance, for the interpretation of wound photographs, color is an indicator of the healing state of the wound and absolute correct representation of colors is important. Dermatologists are calling for standardization.[6] The use of ICC profiles for achieving a good color reproduction across devices[7] is already a well-established practice in different fields including pathology.[8,9] By connecting the acquisition device profile to the display profile, it is possible to achieve good absolute and media relative color reproducibility thanks to colorimetric rendering intents. ICC MIWG is currently working on a definition of the best practices for digital color photography in medicine, which will cover the use cases where high color fidelity is required.

However some modalities use colors to display numerical or quantitative information on top of grayscale images. The exact absolute color used to visualize quantitative information is of less importance, as long as differences are easily perceivable and it is easy for the observer to visually determine what quantitative value is being represented by a specific color. This quantitative imaging approach typically relies on color scales to represent calculated values. Figure 1 is an example of a commonly used "rainbow" color scale, which corresponds with a piecewise linear trajectory in the RGB space. Clearly, equal distances along the color scale do not result in equal perceptual color differences. Also, depending on the exact absolute value of a quantitative value (and the corresponding color), it may be easy or difficult to perceive small differences in that quantitative value.

The display is the component that in the end converts the electric signal to visible light and therefore the choice of color scale and display hardware[2] affects the visual comparative analysis of pseudocolor images.[10] A perceptually linear color scale could help to optimize the visualization of the quantitative colors and reveal subtle details in the image. This can only be accomplished by taking into account the characteristics of the display used for visualization. Using the ICC framework[7] in order to obtain absolute correct reproduction of colors does not solve the problem since the visualized colors will not have equal perceptual separation from each other, and as a consequence, color scales will not be visualized in a perceptually linear way and visualization of subtle color differences will be suboptimal. It theoretically is possible to calibrate a display system such that its entire gamut will correspond to a perfectly perceptually uniform color space[11] [such as, e.g., Lab or ΔE_{2000} (Ref. 12)]. However such calibration will result into an unacceptable reduction of

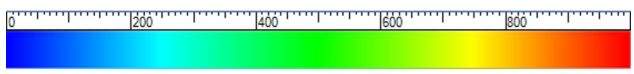

FIG. 1. Commonly used rainbow color scale. (See color online version.)

display color gamut, display luminance, and display contrast and the resulting display would not be DICOM GSDF compliant anymore.

This paper proposes a solution for the above mentioned problem. CSDF is an extension of the DICOM GSDF standard toward color. CSDF defines how color medical displays need to be calibrated and how QA can be performed to obtained consistent, perceptually linear behavior not only for grayscale but also for color and this without compromising display gamut, contrast, or luminance. CSDF is not intended as a solution for true color medical images but is beneficial where quantitative information is being visualized in color. This is the case, for example, for pseudocolors overlaying grayscale images but also modalities where a color image is directly generated (PET, Doppler ultrasound, functional MRI, etc).

3. METHODS

3.A. Proposed CSDF calibration algorithm

DICOM GSDF is a widely adopted standard and has been used for many years with proven performance. Therefore care needs to be taken that perceptually linear calibrating the color behavior of a display does not negatively impact DICOM GSDF calibration quality. The proposed CSDF calibration satisfies that requirement because it is an extension of DICOM GSDF that starts from a DICOM GSDF calibrated display and then additionally calibrates the color behavior of the display system such that not only the grayscale levels but also the color differences show equal perceptual difference.

As mentioned above, the starting point for a CSDF calibration is a DICOM GSDF calibrated display system. This means that the display complies with the GSDF standard but also that other possible corrections that are applied to the grayscale behavior of the display have been applied already. Such corrections typically include, e.g., setting the color point of the neutral gray (e.g., to clearbase or bluebase) and gray tracking [to make sure the color tint of the grayscales is stable as described in AAPM TG196 and IEC62B/MT51 (Ref. 13)]. Calibration to DICOM GSDF also takes into account the ambient light of the viewing environment, by adding the reflected ambient illumination to the luminance values measured from the display, and using these values as a basis for the calibration.

Next a number of measurements need to be done in order to characterize the perceptual linearity of the display system throughout its color gamut. Luminance and color values (Y,x,y) for several color lines are measured throughout the display RGB cube as depicted on Fig. 2 (assuming a display system with 8 bits per color such that the minimum value is 0 and the maximum driving value is 255):

- The neutral gray behavior (from black $[(0,0,0)]_{RGB}$ to white $[(255,255,255)]_{RGB}$).
- The different edges of the RGB cube.
- Sweeps from black $[(0,0,0)]_{RGB}$ to the primary colors (red $[(255,0,0)]_{RGB}$, green $[(0,255,0)]_{RGB}$ and blue $[(0,0,255)]_{RGB}$).

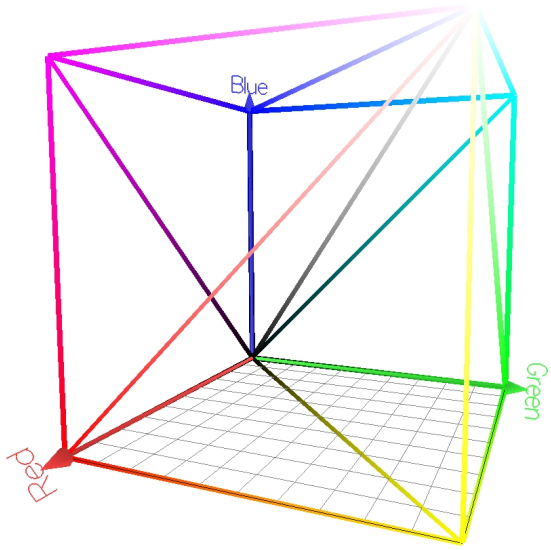

FIG. 2. Measurement trajectories in RGB space for the calculation of the CSDF calibration.

- Sweeps from primary colors (red $[(255,0,0)]_{RGB}$, green $[(0,255,0)]_{RGB}$ and blue $[(0,0,255)]_{RGB}$) to white $[(255,255,255)]_{RGB}$.
- Sweeps from secondary colors (cyan $[(0,255,255)]_{RGB}$, magenta $[(255,0,255)]_{RGB}$ and yellow $[(255,255,0)]_{RGB}$) to black $[(0,0,0)]_{RGB}$.

The contribution of reflected ambient illuminance—both its luminance and chromaticity—should then be added to these measurements. If the chromaticity has not been measured, using the chromaticity of the display's full white may be an acceptable approximation, depending on the ambient viewing conditions. For each of the measured color lines throughout the RGB cube, the perceptual distance ΔE_{2000} (Ref. 12) per step along the curves is calculated. The choice for ΔE_{2000} was made because this is a widely accepted metric that well describes perceptual distance as observed by human observers. Next, for each measured line individually, the points on that line are redistributed on the line such that the perceptual distance ΔE_{2000} between each consecutive pair of points becomes a constant value. As indicated in Fig. 3, this constant ΔE_{2000} step size will still be different from one specific color line to another, because the cumulative ΔE_{2000} values along these lines are different. In other words, one color line (e.g., typically green) may have a larger cumulative perceptual distance along the entire line than another color line (e.g., typically blue). This is shown in Fig. 3. Not enforcing the same absolute perceptual step size on each of the color lines is a very careful decision. If one wants to have the same absolute step size along color lines, then the color contrast of all lines needs to be reduced to that of the line with the lowest average perceptual step size. This would result into a serious reduction in display color gamut, display luminance, and display contrast. Therefore we have chosen to do the perceptual linearization individually per color line and as such the gamut, luminance, and contrast of the display are not reduced.

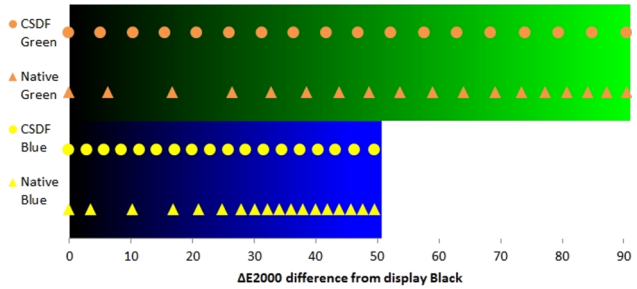

FIG. 3. Calibration concept: redistribution of points on the color lines (e.g., green and blue).

Once all color lines have been processed, then the rest of the color gamut is adapted to ensure a smooth transition from the grayscale line (calibrated to GSDF) to the ΔE_{2000} calibrated color lines. Bilinear interpolation is for example a good choice to perform this smoothing. It is very important to stress that the neutral color gray line (grayscale line) is not adapted. Therefore the DICOM GSDF compliance is not changed because of the additional CSDF calibration step.

3.B. Proposed QA metrics for CSDF

The goal of perceptually linear calibration of a display is that equal differences in RGB drive levels at the display input produce equal perceptual differences in the display output. We suggest using a number of metrics that allow quantifying the degree of perceptual linearity.

In general, perceptual linearity can then be evaluated as follows. First, a set of relevant trajectories (e.g., the lines defined in Sec. 3.A) is defined inside or on the boundary of the RGB color space. Along each line, a series of RGB triplets is chosen, separated by equal increments in the R, G, and/or B values. Next, the display is driven with those RGB values and the display output is measured with a colorimeter and (Y,x,y) values are obtained. The perceptual differences between the measured colors corresponding to consecutive RGB triplets are then calculated as a ΔE_{2000} value. Finally, a metric is defined which expresses to which extent the ΔE_{2000} values differ amongst each other, where a smaller difference corresponds with more perceptual linearity. A specific QA metric proposal for perceptual linearity is described in Sec. 3.B.

Next to perceptual linearity of individual color lines also perceptual smoothness of the entire display is important.[14] Perceptual smoothness of a calibrated display refers to the absence of visually apparent discontinuities when displaying color ramps. In Sec. 3.B, we define a QA metric for perceptual smoothness, based on the methodology proposed by Green for estimating the smoothness of a color transform.[15]

Finally we also propose a visual test pattern that allows for the visual inspection of perceptual linearity of a display system.

3.B.1. Perceptual linearity metric

Along the grayscale, we quantify the perceptual linearity using the traditional DICOM GSDF compliance check, as defined by AAPM TG18,[16] which is based on 18 grayscale measurements. For color, we quantify the perceptual linearity of a display based on the measurements of the display output obtained by sweeping the primaries and secondaries. We define a series of 18 RGB triplets between black and red, with equal steps in the R channel value between them. The choice for 18 values is inspired by the fact that AAPM TG18 also uses 18 measurement values (to evaluate DICOM GSDF compliance along the grayscale). The corresponding 8-bit values are (R,G,B) = (0,0,0), (15,0,0), . . . , (255,0,0). Likewise, a series of 18 RGB triplets are defined between black and the other primaries green and blue, between black and the secondaries cyan, magenta, and yellow, and between black and white (cf. Fig. 4). The corresponding 8-bit values for the sweep from black to yellow, e.g., are (R,G,B) = (0,0,0), (15,15,0), . . . , (255,255,0). Please note that these measured values are all located on the color lines that were measured and used during calibration. Leaving out the duplicates for black results in 120 unique RGB triplets ($7*18-6=120$), for which the corresponding display output is measured as XYZ or Yxy values with a colorimeter. With modern colorimeters the time required to perform these measurements is very reasonable (typically between 1 and 3 min). Measurements corresponding with luminance levels below 5 cd/m^2 are discarded because they fall outside the luminance range of photopic vision (the human visual system has poor discrimination between colors at low luminances). This same luminance threshold is applied in the grayscale chromaticity evaluation procedure defined in an amendment to IEC 62563-1.[13,17] For example, if (R,G,B) = (60,60,60) is the first measurement in the black–white sweep which results in a Y value of at least 5 cd/m^2, then the measurements (R,G,B) = (0,0,0), . . . , (45,0,0) of the black–red sweep are discarded (and similarly for the other sweeps). All nondiscarded measurements are then converted to $L^*a^*b^*$, taking the XYZ of full white as the reference white point. Next, we calculate ΔE_{2000} between the consecutive measurements in each of the six sweeps for the primaries and secondaries, resulting in six series of ΔE_{2000} values. Each

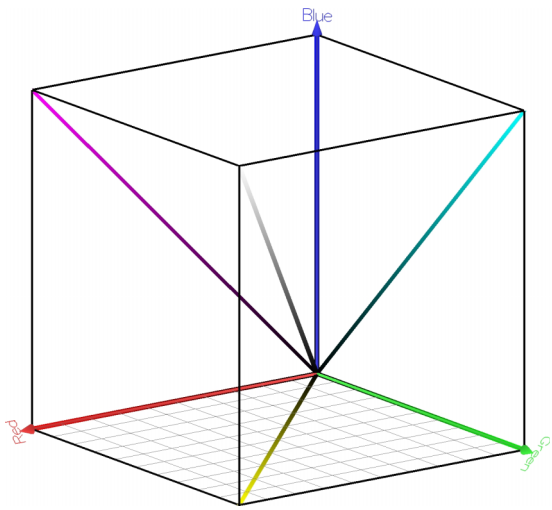

FIG. 4. Measurement trajectories in RGB space for the perceptual linearity metric.

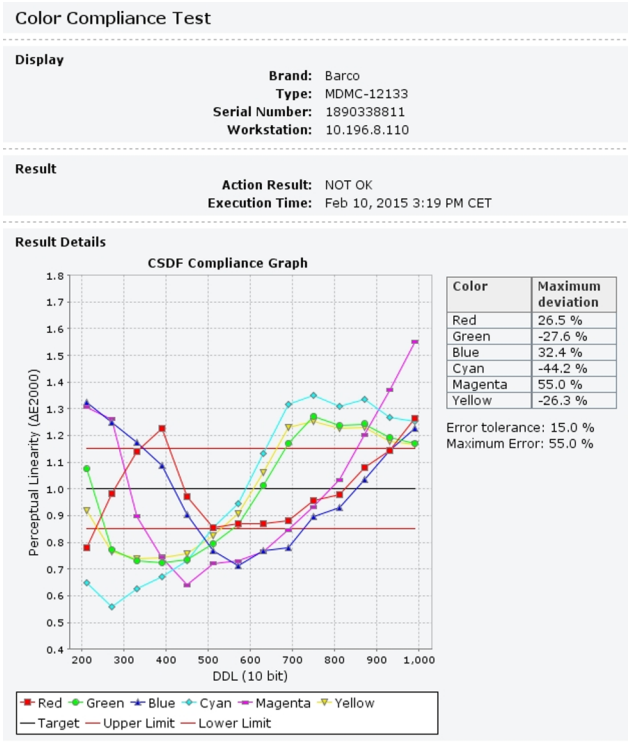

FIG. 5. Perceptual linearity metric for a display without perceptual linearity calibration.

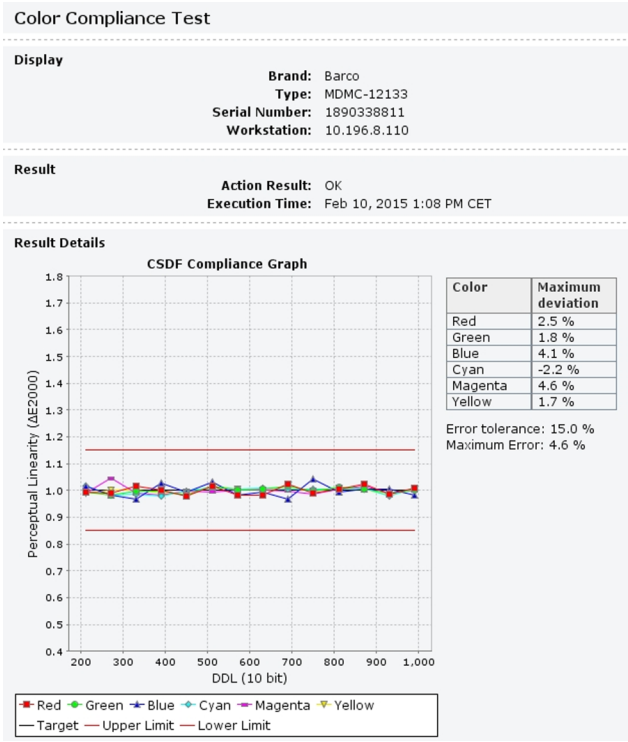

FIG. 6. Perceptual linearity metric for a display with perceptual linearity calibration.

series of ΔE_{2000} values is then normalized by dividing by their average ΔE_{2000} value within the series. For an ideal perceptually linear display, the resulting normalized curves would all be horizontal at exactly 1.0. For each of the six sweeps, we quantify the perceptual linearity as the maximum percentage deviation from unity within the curve. The overall perceptual linearity is quantified as the maximum percentage deviation from 1.0 encountered in any of the six curves. Figures 5 and 6 illustrate the QA metric for perceptual linearity. Figure 5 shows a typical result for a DICOM GSDF calibrated display without calibration for perceptual linearity. The maximum deviation of 55% is caused by a much larger than average ΔE_{2000} between the last two measurements in the magenta sweep. Figure 6 shows a typical result for a DICOM GDSF calibrated display with calibration for perceptual linearity, where the maximum deviation is only 4.6%. If the perceptual linearity metric value is within a predefined tolerance range, the display calibration is considered to be perceptually linear. As an example, a tolerance range of ±15% has been indicated in the Figs. 5 and 6.

3.B.2. Perceptual smoothness metric

In 2008, Green[15] proposed a methodology for estimating the smoothness of a color transform of a ramp, which is represented by Fig. 7. In the context of our paper, the color transform represents the calibrated display. For any input colored ramp consisting of n different colors, ΔE_{2000} is computed between consecutive colors as reproduced by the display, resulting in $n-1$ ΔE_{2000} values. From this resulting ramp, a second derivative is calculated by subtracting two consecutive ΔE_{2000} values, resulting in a set of $n-2$ values, for which the 95th percentile is calculated. This 95th percentile is used to quantify the smoothness of the display's behavior for the given input ramp. Lower values correspond with more smoothness. In order to consider the entire color space, we apply Green's metric on a large number of ramps through the RGB cube: The RGB cube is sampled on a $50\times50\times50$ RGB grid, from which we take a total of 7500 ramps consisting of 50 colors, along the R, G, and B directions. This is illustrated in Fig. 8. The smoothness metric is calculated for each input ramp, and the global average, standard deviation, and maximum values are used as the overall smoothness metric.

3.B.3. Visual test pattern

Figure 9 shows a test pattern that has been defined with the goal of being able to verify visually whether or not a display is perceptually linear. On the right half of the figure, the effect of the CSDF as calculated for a theoretical sRGB display has been applied and exaggerated for comprehension purpose. It is divided into six horizontal areas containing color variations between a primary color (R, G, or B) and a secondary color (C, M, or Y). In the central region of

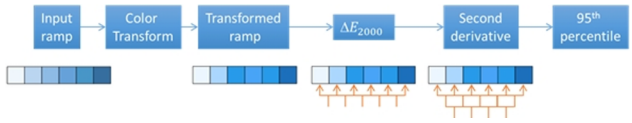

FIG. 7. Perceptual smoothness metric.

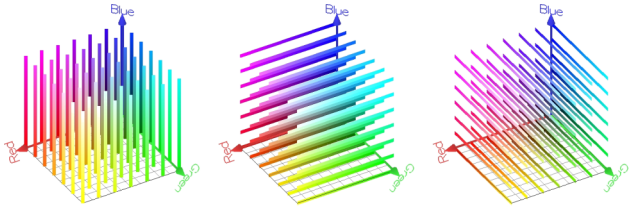

FIG. 8. Color ramps for perceptual smoothness evaluation.

each horizontal area, a "CSDF"-logo in a pure primary color [e.g., (R,G,B) = (255,0,0) in the first row] is embedded in a uniform background color which is slightly different from the primary color [e.g., (R,G,B) = (255,15,0) in the first row, which represents a 5.9% difference in the green channel]. The left and right parts of each horizontal area show linear ramps between the primary and secondary colors, superimposed by a sinusoidal variation in the R, G, or B component. For example, in the top part of the first horizontal area, there is a linear ramp from (R,G,B) = (255,0,0) to (255,255,0) and back to (255,0,0), superimposed by a sinusoidal variation of the green component with an amplitude of 7.65 (3% of 255). Table I lists the amplitudes of the sinusoidal variations and the differences between the logos and their backgrounds in percent. When viewing the pattern on a display which is not perceptually linear, the logos are typically not visible, and the sinusoidal variations in the color ramps are typically not visible throughout the ramp but rather in certain parts of the ramp. On a display with perceptually linear calibration, however, the logos clearly stand out against their backgrounds, and the sinusoidal variations are visible throughout the ramp.

4. RESULTS

The validation of this method via observer studies will be the topic of another paper yet to come. Results presented here are a summary of our preliminary observations.

The calibration method presented above has been evaluated on a wide variety of displays using the metrics presented in Sec. 3.B.1. Those displays can be arranged in three categories:

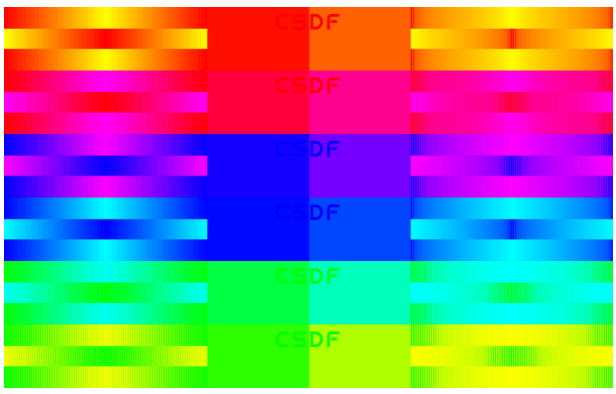

FIG. 9. Visual test pattern for perceptual linearity. The left side shows the test pattern without modifications. The right part is a representation of the effect of the CSDF calibration. This effect has been amplified to make it more visible.

TABLE I. Sinusoidal amplitudes and logo-background differences in the visual test pattern.

	Amplitude (%)	Difference (%)
Red–yellow	3.0	5.9
Red–magenta	15	24
Blue–magenta	10	7.8
Blue–cyan	3.0	3.5
Green–cyan	20	26
Green–yellow	25	17

- Consumer displays are most commonly available in the market. They typically have a gamut that resembles sRGB and have a relatively low brightness.
- Professional displays are high quality monitors, superior to consumer displays in term of brightness. They often also integrate some internal sensors to ensure color and brightness stability. The distinction between professional and consumer displays is not related to the market they target but to their specifications.
- Medical displays are professional displays with a DICOM GSDF calibration instead of the usual sRGB model (which almost follows a gamma 2.2) and they typically have even more built-in electronics to ensure stable luminance and color behavior.

All measurements have all been done using a Konica Minolta CA-310. A warm-up period of 2 h has been respected before proceeding to any measurements for calibration or validation to ensure a stable display behavior.

4.A. Bench testing results

4.A.1. *Consumer displays*

Dell U2410 is a typical consumer display covering the sRGB gamut. The characteristics observed during these tests are given in Table II.

Figure 10 shows the perceptual linearity of the calibrated display. The overall deviation is 12.1%.

Dell U2713H is a LCD display with a LED backlight. With an specified brightness of 350 cd/m^2 and a gamut covering 99% of Adobe RGB, this display presents high end characteristics. However this display does not contain any sensor for self-monitoring, and for this reason, it is considered

TABLE II. Measured characteristics of the Dell U2410 used to test the calibration.

	Y (cd/m^2)	x	y
Black	0.2515	0.274	0.2832
Red	52.38	0.6448	0.3356
Green	164.9	0.3289	0.5814
Blue	14.9	0.1445	0.0568
White	228.0	0.3267	0.3305
Contrast		907:1	
Display function		Gamma 2.2	

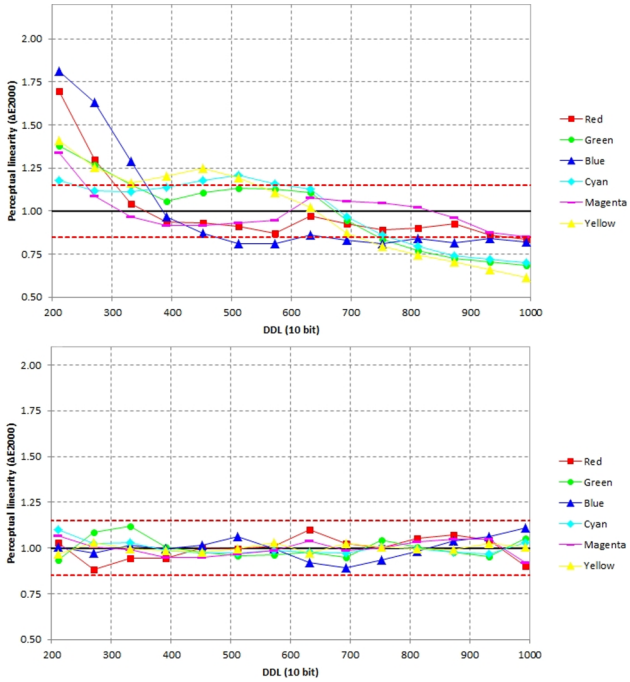

FIG. 10. Perceptual linearity of a calibrated Dell U2410 without (above) and with (below) CSDF.

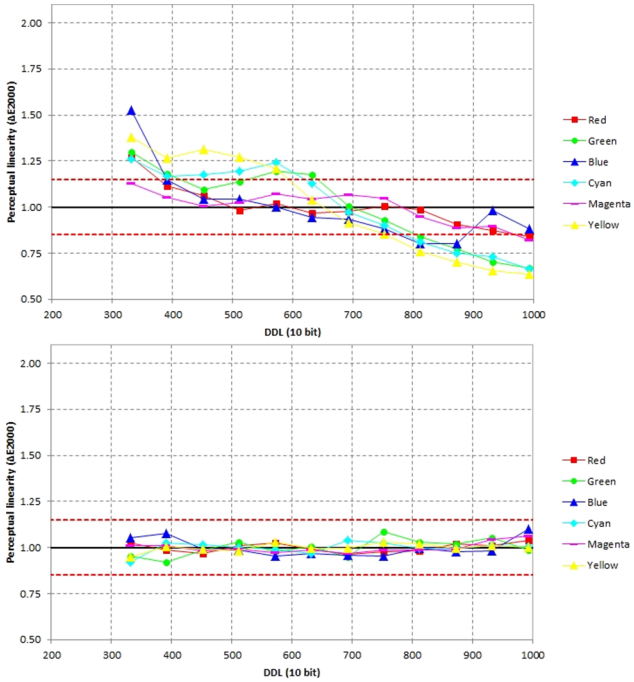

FIG. 11. Perceptual linearity of a typical consumer display (Dell U2713H) without (above) and with (below) CSDF. Please note that because of the relatively low luminance of this display, more measurements are discarded because they are below the threshold of 5 cd/m^2.

as a consumer display. The characteristics observed while running the tests presented here are summarized in Table III.

Figure 11, representing the perceptual linearity on the calibrated display, shows an overall deviation of 10.27% while it is 98.12% without CSDF.

4.A.2. Professional displays

Barco MDCC-6330 (adapted to gamma 2.2*)* is in fact a high end medical display. But the display has been set to a gamma 2.2 display function, a *D*65 white color and a brightness of 600 cd/m^2 in order to resemble the characteristics of a typical professional display. The characteristics observed while running the tests presented here are summarized in Table IV. Figure 12 shows the perceptual linearity of the display without calibration. The overall deviation is 113.62%. The highest deviations are observed on the very first levels of blue, red, and magenta sweeps. These are the darkest colors of the control set. The luminance of those colors is below 3 cd/m^2. Once calibrated, the deviation is reduced to 10.86%.

4.A.3. Medical displays

Barco MDCC-6330 (GSDF) is this time used as a medical display by setting it to DICOM GSDF display function, *D*65 white color, and a brightness of 600 cd/m^2. Figure 13 shows the perceptual linearity of the display without calibration. The overall deviation is equal to 163.5%. This value is lower than when using the display as a professional one because DICOM GSDF tends to reduce the luminance of the lowest levels and then reduce the difference between the control points that were critical with gamma 2.2 setting. The calibration reduces the deviation to 10.84% as illustrated by Fig. 13. This tends to demonstrate that the accuracy of the calibration depends more on the shape of the display gamut than on the repartition of the color into this gamut.

Barco MDPC-4130 is a medical display embedding a monitoring system and an automatic DICOM GSDF calibration system. The characteristics observed while running the

TABLE III. Measured characteristics of the Dell U2713H used to test the calibration.

	Y (cd/m^2)	x	y
Black	0.20	0.2581	0.3038
Red	60.95	0.6356	0.3368
Green	118.5	0.2013	0.761
Blue	14.26	0.1471	0.0617
White	195.7	0.3162	0.3433
Contrast		979:1	
Display function		Gamma 2.2	

TABLE IV. Measured characteristics of the Barco MDCC-6330 used to test the calibration.

	Y (cd/m^2)	x	y
Black	0.5942	0.2362	0.2466
Red	99.03	0.6476	0.3315
Green	573.5	0.312	0.6165
Blue	77.79	0.144	0.0655
White	738.0	0.2707	0.3101
Contrast		1242:1	
Display function		Gamma 2.2	

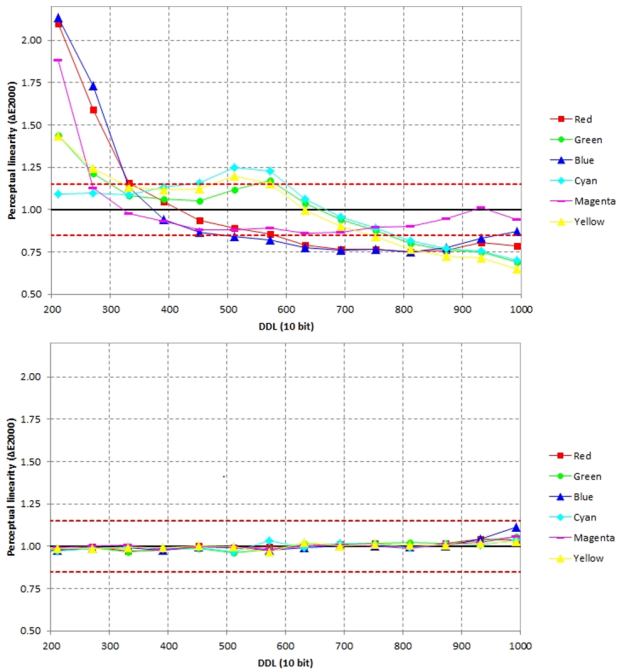

FIG. 12. Perceptual linearity of a typical professional display (modified Barco MDCC-6330) without (above) and with (below) CSDF.

tests presented here are summarized in Table V. The CSDF calibration has been applied on top of GSDF. Figure 14 shows the effect of GSDF on the perceptual linearity of a display. The measured overall uniformity is equal to 50.61%. CSDF is able to make these color sweeps much more perceptually linear as shown at the bottom of Fig. 14. Perceptual linearity deviation is now equal to 3.89%.

TABLE V. Measured characteristics of the Barco MDPC-4130 used to test the calibration.

	Y (cd/m^2)	x	y
Black	0.4779	0.2563	0.2562
Red	57.88	0.6417	0.3229
Green	281.4	0.3107	0.6110
Blue	47.1915	0.1547	0.0731
White	387.1	0.2786	0.3015
Contrast		810:1	
Display function		Gamma 2.2	

4.B. Simulation results

Perceptual smoothness has been evaluated based on a simulated display with native characteristics as described in Table VI. This display has been calibrated to DICOM GSDF and to CSDF. In both cases, perceptual smoothness has been evaluated according to the method described in Sec. 3.B.2, based on simulated measurements on the calibrated display. The results are shown in Table VII. The average, standard deviation, and maximum value of the smoothness metric show an improvement (lower values are better) of perceptual smoothness for CSDF compared to DICOM GSDF.

4.C. Visual results

The quality of the CSDF calibration can be visually observed directly on the different displays systems. The test pattern described in Sec. 3.B.3 and presented on Fig. 9 is displayed on the different monitors presented above with and without the proposed calibration to visualize the difference of perception of colors. All figures shown in this section are

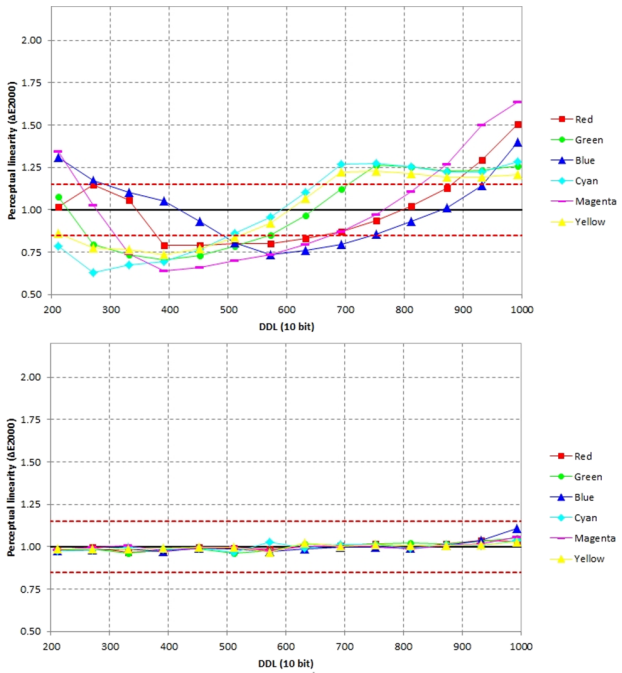

FIG. 13. Perceptual linearity of a typical medical display (Barco MDCC-6330) without (above) and with (below) CSDF.

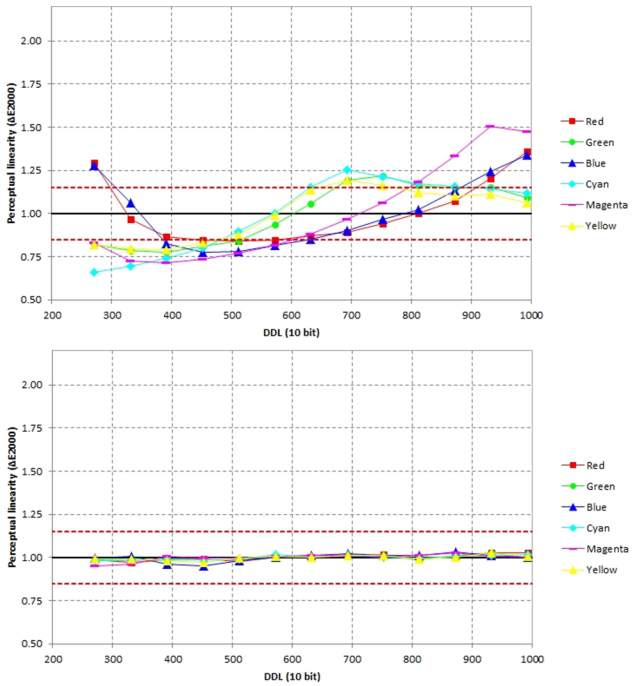

FIG. 14. Perceptual linearity of a typical medical display (Barco MDPC-4130) without (above) and with (below) CSDF.

TABLE VI. Native characteristics of the simulated display used for perceptual smoothness evaluation.

	Y (cd/m^2)	x	y
Black	0.60	0.3127	0.3290
Red	128.08	0.6400	0.3300
Green	429.26	0.3000	0.6000
Blue	43.86	0.1500	0.0600
White	600.0	0.3127	0.3290
Contrast		1000:1	
Display function		sRGB	

photographs that were taken (with a standard digital camera) of the tested display systems.

4.C.1. Consumer displays

Figure 15 shows the Dell U2410 consumer display with and without CSDF calibration applied. On this monitor, CSDF calibration makes the low contrast text more readable in the middle of the image.

Figure 16 focuses on the sinusoidal part of the test pattern. It shows that CSDF makes more bands visible on the different color ramps, proving that it helps in improving visualization of subtle color differences. The same conclusion can be drawn for professional displays (see Fig. 18) and medical displays (see Fig. 20).

4.C.2. Professional displays

Figure 17 presents a professional display with a CSDF calibration applied and without the color calibration. The noncalibrated display is set to Gamma 2.2, and its white point is *D*65. Here again, CSDF helps perceiving the low contrast targets in the middle on the pattern, even if the effects seem to be less pronounced here than on the tested consumer monitors. Figure 18 also shows the improved color differentiation on high frequency signals. For clarity reasons, the circular regions of Fig. 18 show a magnified area where differences are particularly well visible.

4.C.3. Medical displays

In Fig. 19, one can observe that the low contrast image in the central column of the pattern is much more visible with CSDF calibration than with DICOM GSDF, especially in greenish and blueish areas. Figure 20 focuses on the sinusoidal part of the test pattern. For clarity reasons, the circular regions of Fig. 20 show a magnified area where differences are particularly well visible. It is to be noted that some bands are more visible with GSDF compared to CSDF, but there are also color values for which there are no bands visible anymore in the GSDF mode. This is because for specific colors the ΔE_{2000} per step value in the GSDF mode sometimes is larger but in other cases much lower than in the CSDF mode. However after CSDF calibration the result is much more consistent and the total number of distinguishable bands is higher.

TABLE VII. Perceptual smoothness results.

Calibration	Average	Std. dev.	Maximum
GSDF	0.2202	0.1522	1.3098
CSDF	0.1955	0.1250	1.1279

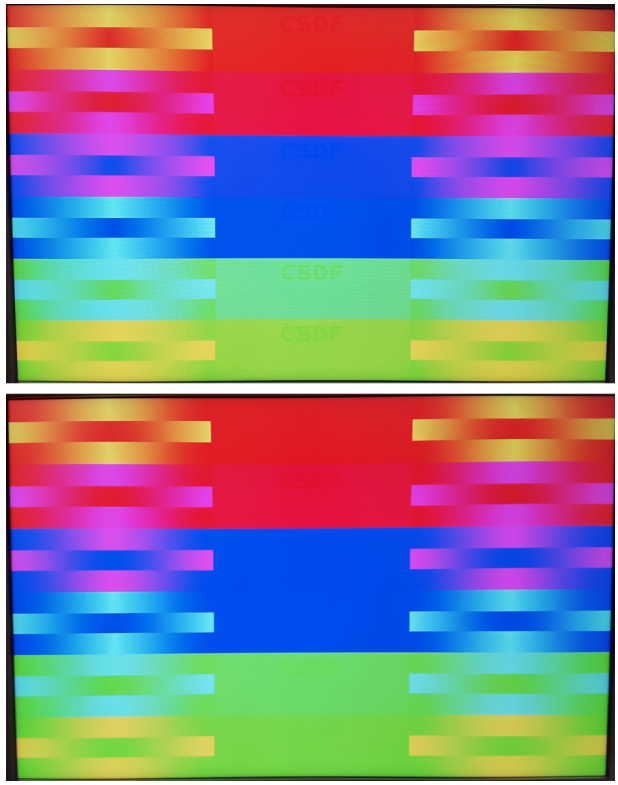

FIG. 15. Photograph of the test pattern seen on a consumer display (Dell U2410) with CSDF applied (above) and without (below). (See color online version.)

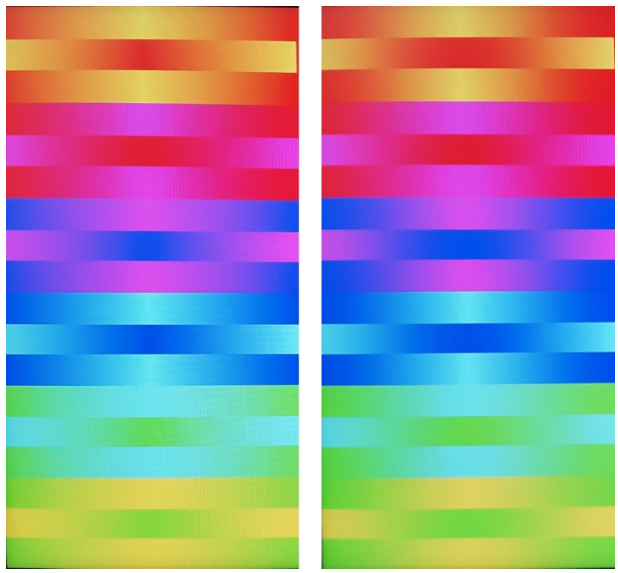

FIG. 16. Zoom on the sinusoid part of the test pattern seen on a consumer display (Dell U2410) with CSDF applied (left) and without (right). (See color online version.)

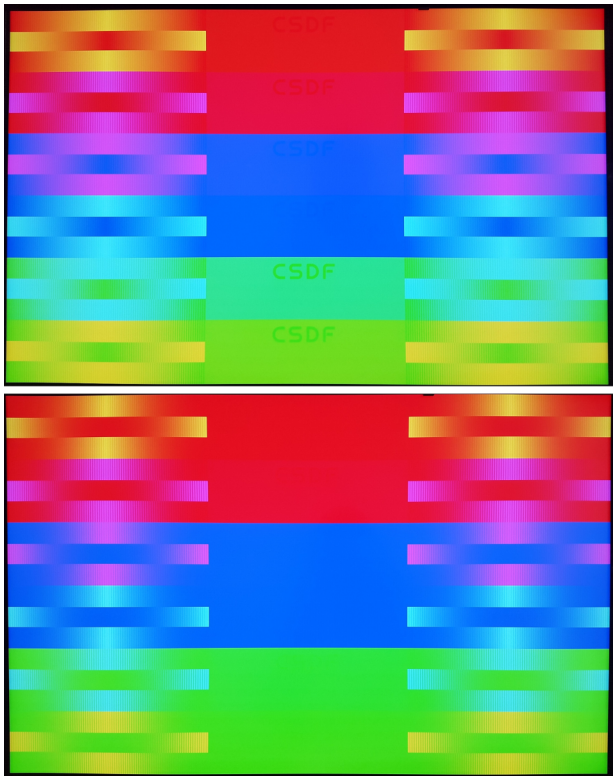

FIG. 17. Photograph of the test pattern seen on a professional display (modified Barco MDCC6330 without DICOM GSDF) with CSDF applied (above) and without (below).

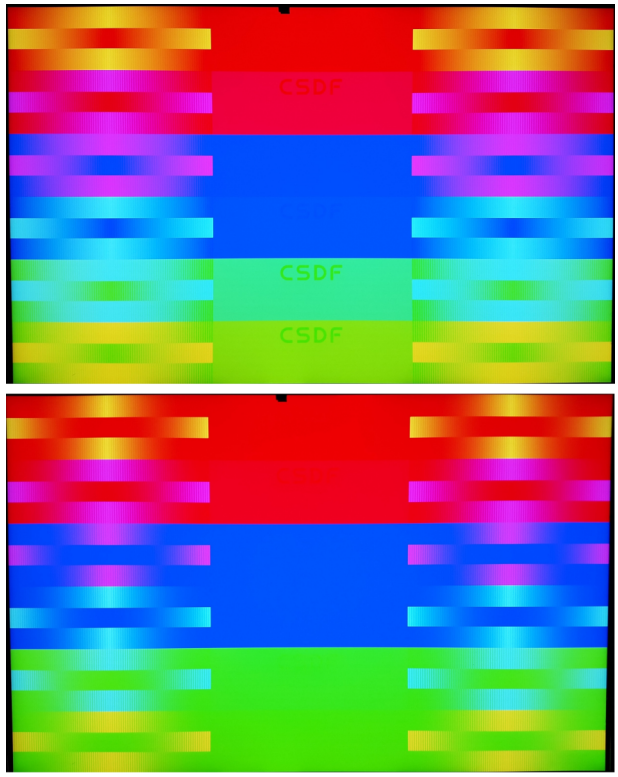

FIG. 19. Photograph of the test pattern seen on a medical display (Barco MDPC4130 set to DICOM GSDF) with CSDF applied (above) and without (below). (See color online version.)

5. DISCUSSION

5.A. Application domains

As already mentioned in the problem statement, CSDF is intended for applications where quantitative information is being visualized in color. It is not intended as a solution for true color medical images which may require faithful absolute color reproduction. However, CSDF does not preclude absolute color reproduction. The behavior of a CSDF calibrated display can be described in detail in an ICC profile, which can be used by a color management framework in the

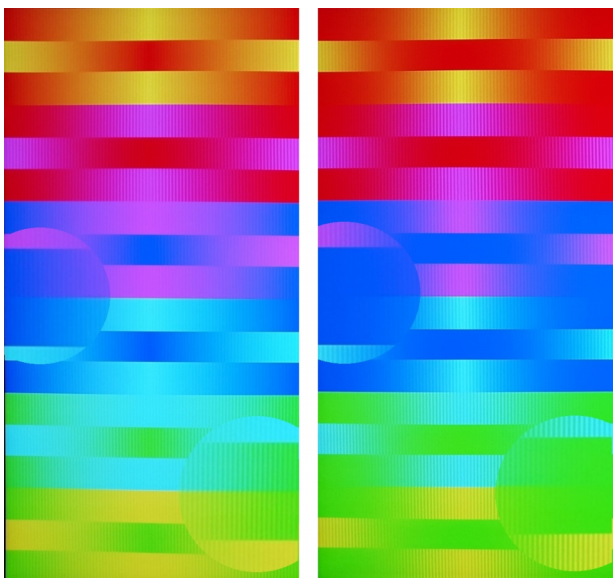

FIG. 18. Zoom on the sinusoid part of the test pattern seen on a professional display (Barco MDCC6330 without DICOM GSDF) with CSDF applied (left) and without (right).

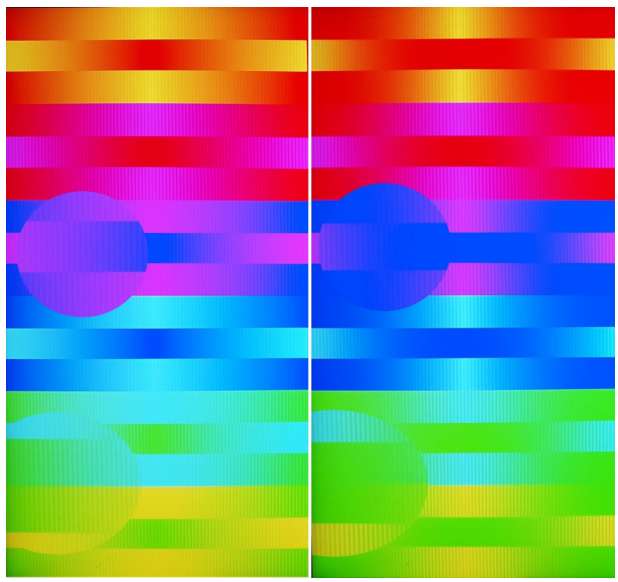

FIG. 20. Zoom on the sinusoid part of the test pattern shown on a medical display (Barco MDPC4130 set to DICOM GSDF) with CSDF applied (left) and without (right). (See color online version.)

viewing application or operating system, to obtain absolute color reproduction on the CSDF calibrated display.

5.B. Evaluation of the CSDF calibration and QA algorithms

The proposed CSDF calibration and quality assurance algorithms have been evaluated on a wide variety of display systems. The results show that the proposed calibration algorithm works well both on consumer displays as well as on professional displays and medical displays. The proposed QA metrics show that for all tested displays the resulting error compared to the CSDF target can be reduced (from 98% to 164%) to well beyond the 15% tolerance level.

The visual test pattern also is capable of visualizing the effect of CSDF calibration. Although the visual pattern itself is not sufficient to determine whether a display is within 15% of the CSDF target, it is possible to observe that the CSDF calibration improves perceptual linearity and ensures that more subtle color differences are visible.

5.C. Standardization and future work

It is our intention to achieve standardization of the proposed CSDF calibration and QA algorithms. A first step in that direction has already been taken. In context of the ICC MIWG Displays,[18] a white paper has been created and approved that describes how to visualize color medical images in a ICC framework context.

Furthermore also contact has been made with AAPM TG196 and DICOM WG11 to determine what standardization committee is the best fit, and how a task can be started.

6. CONCLUSIONS

The use of color in medical imaging has become standard practice rather than exception. Color is not used anymore solely for annotations and metadata but also as a basis for primary diagnosis. Unfortunately as of today, there are no agreed upon standards how color in medical imaging needs to be visualized.

This paper proposed CSDF as a possible extension of the DICOM GSDF calibration standard. The goal of CSDF is to ensure consistent, perceptually linear representation of both grayscale and color information. However, this calibration method may not be necessary for modalities requiring an exact color reproduction.

The proposed CSDF algorithm was described in detail and has been validated on a range of display systems from consumer level to high-end medical displays. QA metrics were proposed to quantitatively measure calibration compliance. Preliminary results show that the algorithm is stable and largely improves perceptual linearity on all tested displays. A deeper study of this calibration involving both objective and subjective validations on a larger scale will be presented in a future publication. Actions have started for working toward standardization of CSDF as a generally accepted international standard.

ACKNOWLEDGMENT

This work was supported by the Flemish Government through the "Agentschap Innoveren en Ondernemen (IWT Grant No. 130472)."

CONFLICT OF INTEREST DISCLOSURE

The authors of this paper are employees of Barco NV, a company designing and manufacturing medical display systems. Barco NV is openly working toward international standardization of the calibration and QA methods proposed in this paper.

[a]Electronic mail: tom.kimpe@barco.com

[1]A. Badano, C. Revie, A. Casertano, W.-C. Cheng, P. Green, T. Kimpe, E. Krupinski, C. Sisson, S. Skrøvseth, D. Treanor, P. Boynton, D. Clunie, M. Flynn, T. Heki, S. Hewitt, H. Homma, A. Masia, T. Matsui, B. Nagy, M. Nishibori, J. Penczek, T. Schopf, Y. Yagi, and H. Yokoi, "Consistency and standardization of color in medical imaging: A consensus report," J. Digital Imaging **28**, 41–52 (2015).

[2]T. Kimpe, J. Rostang, A. Avanaki, K. Espig, A. Xthona, I. Cocuranu, A. Parwani, and L. Pantanowitz, "Does the choice of display system influence perception and visibility of clinically relevant features in digital pathology images?," Proc. SPIE **9041**, 904109 (2014).

[3]P. G. Barten, "Formula for the contrast sensitivity of the human eye," Proc. SPIE **5294**, 231–238 (2003).

[4]K. A. Fetterly, H. R. Blume, M. J. Flynn, and E. Samei, "Introduction to grayscale calibration and related aspects of medical imaging grade liquid crystal displays," J. Digital Imaging **21**, 193–207 (2008).

[5]J. A. Carrino, "Digital image quality: A clinical perspective," in *Quality Assurance* (The Society for Computer Applications in Radiology, Great Falls, VA, 2003), pp. 29–37.

[6]A. A. Marghoob, "Standards in dermatologic imaging," JAMA Dermatol. **151**(8), 819–821 (2015).

[7]L. D. Silverstein, S. F. Hashmi, K. Lang, and E. A. Krupinski, "Paradigm for achieving color-reproduction accuracy in LCDs for medical imaging," J. Soc. Inf. Disp. **20**, 53–62 (2012).

[8]W. C. Revie, M. Shires, P. Jackson, D. Brettle, R. Cochrane, and D. Treanor, "Color management in digital pathology," Anal. Cell. Pathol. (Amsterdam) **2014**, 652757.

[9]E. A. Krupinski, L. D. Silverstein, S. F. Hashmi, A. R. Graham, R. S. Weinstein, and H. Roehrig, "Impact of color calibration on breast biopsy whole slide image interpretation accuracy and efficiency," in *Breast Imaging* (Springer, Switzerland, 2014), pp. 744–748.

[10]S. Zabala-Travers, M. Choi, W.-C. Cheng, and A. Badano, "Effect of color visualization and display hardware on the visual assessment of pseudocolor medical images," Med. Phys. **42**, 2942–2954 (2015).

[11]I. Lissner and P. Urban, "Toward a unified color space for perception-based image processing," IEEE Trans. Image Process. **21**, 1153–1168 (2012).

[12]G. Sharma, W. Wu, and E. N. Dalal, "The CIEDE2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations," Color Res. Appl. **30**, 21–30 (2005).

[13]IEC, Medical electrical equipment–medical image display systems. Part 1. Evaluation methods, IEC 62563-1:2009+AMD1:2016 CSV, 2016.

[14]A. Avanaki, K. Espig, T. Kimpe, A. Xthona, C. Marchessoux, J. Rostang, and B. Piepers, "Perceptual uniformity of commonly used color spaces," Proc. SPIE **9041**, 90410V (2014).

[15]P. Green, "A smoothness metric for colour transforms," Proc. SPIE **6807**, 68070I (2008).

[16]E. Samei, A. Badano, D. Chakraborty, K. Compton, C. Cornelius, K. Corrigan, M. Flynn, B. Hemminger, N. Hangiandreou, J. Johnson, D. Moxley-Stevens, W. Pavlicek, H. Roehrig, L. Rutz, E. Samei, S. Shepard, R. Uzenoff, J. Wang, and C. Willis, "Assessment of display performance for medical imaging systems: Executive summary of AAPM TG18 report," Med. Phys. **32**, 1205–1225 (2005).

[17]IEC, Medical electrical equipment, medical image display systems. Part 1. Evaluation methods, IEC 62563-1:2009, 2009.

[18]T. Kimpe and M. Flynn, "ICC MIWG displays: Draft recommendations for a color visualization pipeline," Med. Phys. **42**, 3670 (2015).